PTO/SB/21 (09-04)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission: 29

Application Number	09/884,403
Filing Date	06/19/01
First Named Inventor	Shamim A. Alpha
Art Unit	2854
Examiner Name	Lamont M. Spooner
Attorney Docket Number	27252.2

ENCLOSURES (Check all that apply)

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD. Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [x] Appeal Communication to TC (Appeal Notice, Brief, Reply Brief)
- [ ] Proprietary Information
- [ ] Status Letter
- [ ] Other Enclosure(s) (please identify below):

Return Post Card

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	McDonald Hopkins Co., LPA		
Signature	Petar Kraguljac		
Printed name	Petar Kraguljac		
Date	8/14/06	Reg. No.	38,520

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below:

Signature	Doreen Zabinski		
Typed or printed name	Doreen M. Zabinski	Date	8/14/06

PTO/SB/21 (09-04)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission: 29

Application Number	09/884,403
Filing Date	06/19/01
First Named Inventor	Shamim A. Alpha
Art Unit	2654
Examiner Name	Lamont M. Spooner
Attorney Docket Number	27252.2

RECEIVED
CENTRAL FAX CENTER
AUG 14 2006

ENCLOSURES *(Check all that apply)*

- ☑ Fee Transmittal Form
  - ☐ Fee Attached
- ☐ Amendment/Reply
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Reply to Missing Parts/ Incomplete Application
  - ☐ Reply to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____
  - ☐ Landscape Table on CD
- ☐ After Allowance Communication to TC
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☑ Appeal Communication to TC (Appeal Notice, Brief, Reply Brief)
- ☐ Proprietary Information
- ☐ Status Letter
- ☐ Other Enclosure(s) (please identify below):

Return Post Card

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	McDonald Hopkins Co., LPA		
Signature	Petar Kraguljac		
Printed name	Petar Kraguljac		
Date	8/14/06	Reg. No.	38,520

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below:

Signature	Doreen Zabinski		
Typed or printed name	Doreen M. Zabinski	Date	8/14/06

This collection of information is required by 37 CFR 1.5. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and1.14. This collection is estimated to 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/SB/17 (12-04v2)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Effective on 12/08/2004.
Fees pursuant to the Consolidated Appropriations Act, 2005 (H.R. 4818).

FEE TRANSMITTAL
For FY 2005

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) 500.00

Complete If Known	
Application Number	09/884,403
Filing Date	06/19/01
First Named Inventor	Shamim A. Alpha
Examiner Name	Lamont M. Spooner
Art Unit	2654
Attorney Docket No.	27252.2

METHOD OF PAYMENT (check all that apply)

☐ Check ☐ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☑ Deposit Account Deposit Account Number: 150635 Deposit Account Name: _____

For the above-identified deposit account, the Director is hereby authorized to: (check all that apply)

☑ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, except for the filing fee

☑ Charge any additional fee(s) or underpayments of fee(s) under 37 CFR 1.16 and 1.17 ☑ Credit any overpayments

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES

Application Type	FILING FEES Fee ($)	FILING FEES Small Entity Fee ($)	SEARCH FEES Fee ($)	SEARCH FEES Small Entity Fee ($)	EXAMINATION FEES Fee ($)	EXAMINATION FEES Small Entity Fee ($)	Fees Paid ($)
Utility	300	150	500	250	200	100	
Design	200	100	100	50	130	65	
Plant	200	100	300	150	160	80	
Reissue	300	150	500	250	600	300	
Provisional	200	100	0	0	0	0	

2. EXCESS CLAIM FEES

Fee Description	Fee ($)	Small Entity Fee ($)
Each claim over 20 (including Reissues)	50	25
Each independent claim over 3 (including Reissues)	200	100
Multiple dependent claims	360	180

Total Claims	Extra Claims	Fee ($)	Fee Paid ($)
19 - 20 or HP =	0	x 50	= 0

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims	Extra Claims	Fee ($)	Fee Paid ($)
3 - 3 or HP =	0	x 200	= 0

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims Fee ($)	Fee Paid ($)

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $250 ($125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	Extra Sheets	Number of each additional 50 or fraction thereof	Fee ($)	Fee Paid ($)
- 100 =	/ 50 =	(round up to a whole number)	x	=

4. OTHER FEE(S)

	Fees Paid ($)
Non-English Specification, $130 fee (no small entity discount)	
Other (e.g., late filing surcharge): Fee for Appeal Brief	500.00

SUBMITTED BY

Signature	Petar Kraguljac	Registration No. (Attorney/Agent) 38,520	Telephone (216) 348-5843
Name (Print/Type)	Petar Kraguljac		Date 8/14/06

RECEIVED
CENTRAL FAX CENTER

AUG 14 2006

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Shamim A. Alpha	Examiner: Lamont M. Spooner Art Unit: 2654
Serial No.: 09/884,403	
Filed: June 19, 2001	
For: METHODS AND SYSTEMS FOR DETERMINING A LANGUAGE OF A DOCUMENT (as amended)	
Date of Final Office Action: January 30, 2006	Attorney Docket No.: 27252.2
Notice of Appeal Filed: June 14, 2006	
Today's Date: August 14, 2006	

APPEAL BRIEF

Mail Stop Appeal Brief – Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

This Appeal Brief is timely provided to support the Notice of Appeal filed June 14, 2006.

CERTIFICATE OF FACSIMILE

Date of Deposit: August 14, 2006

I hereby certify that these papers are being transmitted to The United States Patent and Trademark Office facsimile number (571) 273-8300 on August 14, 2006.

Doreen Zabinski

OID-2000-150-01

08/15/2006 MBINAS 00000024 150635 09884403
01 FC:1402 500.00 DA

1. Real Party in Interest:

The real party in interest is Oracle International Corporation.

The inventor is Shamim A. Alpha, who on June 17, 2001 assigned his interest to Oracle Corporation, a Delaware Corporation with a place of business at 500 Oracle Parkway, Redwood Shores, California, 94065. On October 30, 2003, Oracle Corporation then assigned its interest to Oracle International Corporation (OIC), a California Corporation with a place of business at 500 Oracle Parkway, Redwood Shores, California, 94065. This has been recorded by the U.S. Patent Office on 12/08/2003 at Reel/Frame: 014773/0488.

2. Related Appeals and Interferences

There are no other prior and/or pending appeals, interferences, or judicial proceedings that are related to, directly affect, or that will be directly affected by or have a bearing on the Board's decision.

3. Status of Claims

Claims 1, and 4-21 are pending in the application.
Claims 1, and 4-21 stand rejected.
Claims 2 and 3 have been canceled.
The rejections of claims 1, and 4-21 are appealed.

4. Status of Amendments

No Amendments were filed subsequent to the Final Office Action.

5. Summary of Claimed Subject Matter

The claimed subject matter concerns systems and methods for identifying the language in which a document is written. The systems and methods extract words from the document and update negative assumptions and/or null hypotheses about candidate languages not being the language in which the document is written. The updates are based on probabilities associated with the term. Rather than adding up probabilities that a term is from a certain language for each of the candidate languages and then picking the highest scoring language, the systems and methods instead seek to select a language based on a proof and/or to de-select a language based on a proof (see specification [0026]). The proofs are possible because the probabilities consider the term in the context of all the candidate languages, not just within each individual candidate language (see specification [0027]).

Independent Claim 1

Claim 1 concerns a system for determining the language of a document. Claim 1 is described generally in paragraphs [0006] and [0025-0026] and with respect to Figure 1. Pinpoint citations to specific elements are provided herein. The determination is based on probabilities associated with terms in the document. Claim 1 includes a logic (Fig. 1 120, paragraph [0026]) for setting a negative assumption for a language, the logic establishes a value that facilitates proving that a document is or is not in a certain candidate language. This facilitates pruning a problem space, producing a more efficient algorithm. Additionally, claim 1 describes that a probability associated with a term is based, at least in part, on the occurrence of the term in all candidate languages, not in each candidate language individually (paragraph [0046]). Having the probability (Fig. 1 125, paragraph [0027]) depend on occurrences in all the candidate languages facilitates the pruning since it enables proving a negative assumption. Having the probability for a term depend on occurrences across multiple languages creates a complex non-independent (e.g., related) statistical analysis that allows a negative to be proven.

In claim 1 a system that may start with one hundred candidate languages may quickly narrow down the problem space to a handful of languages based on satisfying a negative assumption using a contrary probability associated with an encountered term.

No such similar pruning is possible in the referenced accumulator based approaches. Although the claim may not use the term "pruning", the action flows directly from the claimed elements and their actions. Unlike simple accumulator based approaches that add independent event term probabilities derived from independent language probability models, the claimed system is interested in eliminating bad candidates, rather than just determining which is the highest scoring candidate. Thus, the claimed system is from a class of algorithms that are fundamentally different from conventional "rank all the contender" algorithms.

Independent Claim 7

Claim 7 describes a method for determining the language of a document based on probabilities associated with terms in the document. Claim 7 is described generally in paragraphs [0008] and [0035-0046] and with respect to Figure 4. Pinpoint citations to specific elements are provided herein. The probabilities associated with terms are based on occurrences of the terms in all candidate languages. The probabilities being all candidate language based facilitates contrary probability processing and null hypothesis processing. Thus, the method includes setting a null hypothesis (Fig. 4 405, paragraph [0036]) for a language. If during processing this null hypothesis is disproved then the language can be selected as the language of the document. Using null hypothesis analysis, one individual language can be selected, rather than a conventional ranking of contenders. Additionally, claim 7 recites determining a contrary probability for a candidate language. The contrary probability facilitates doing more than simply adding probabilities to an accumulator and choosing a relatively higher ranked "winner".

Dependent Claim 10

Claim 10 depends from claim 7 and recites pregenerating probability data corresponding to each candidate language, the probability data including a probability value for a text string that is normalized based on an occurrence probability of the text string in all the candidate languages. Paragraph [0027] and Figure 2 shows a system and

process for generating probability data. Paragraphs [0030-0031] describe normalizing the data based on occurrence values of all selected languages and not isolated views of only one language.

Dependent Claim 14

Claim 14 depends from claim 7 and recites language relating to normalizing occurrences from all the candidate languages. The discussion of normalizing under claim 10 above applies here and the same references to the specification may be used.

Independent Claim 15

Claim 15 concerns a process for determining that a document is written in a selected language. Claim 15 is described generally in paragraphs [0009] and [0036-0043] and with respect to figure 4. Pinpoint citations to specific elements are provided herein. The process includes setting a probability assumption that indicates that the document is not written in the selected language. (Fig. 4 405, paragraph [0036]) Additionally, claim 15 describes the process including disproving the probability assumption based on a contrary probability. (paragraph [0041]) The determining and disproving produce a "last man standing" algorithm where candidates can be eliminated and/or selected. This type of algorithm differs from those where no contrary probabilities are used, like the referenced accumulator based probabilities that rely only on positive probabilities. The contrary probabilities are available in part due to the properties of the probabilities.

Dependent Claims 17-19

Claim 17 depends from claim 16 and claims 18-19 depend from claim 17. Claim 17 recites generating a probability database having a contrary probability where the contrary probability of a character string in one language is determined based on an occurrence frequency of the character string in the one language influenced by a total occurrence frequency of the character string in all the candidate languages.

Paragraph [0027] and Figure 2 shows a system and process for generating probability data. Paragraphs [0030-0031] describe determining occurrence frequencies (claim 18) and normalizing probabilities by the total occurrence frequency of the character string in all candidate languages (claim 19).

Independent Claim 21

Claim 21 concerns a computer program product configured to perform the process claimed in claim 15. The computer program product is described generally in the definition of a computer readable medium in paragraphs [0009], [0020], and [0036-0043].

6. Grounds of Rejection to be Reviewed on Appeal

The following grounds of rejection are to be reviewed on appeal:

I. Claims 1, and 4-13 were rejected under 35 U.S.C. §102(e) as being anticipated by Elworthy (US 6,125,362) (Elworthy).

II. Claims 15, 16, 20, and 21 were rejected under 35 U.S.C. §102(b) as being anticipated by Pon et al. (US 6,047,251) (Pon).

III. Claims 14, and 17-19 were rejected under 35 U.S.C. §103(a) as being unpatentable over Pon in view of Elworthy.

IV. MPEP §2141.03 requires that Office Actions ascertain and describe the level of the hypothetical person of ordinary skill in the art so that objectivity can be maintained. Here the Office Actions neither ascertained nor reported on the level of ordinary skill in the art and thus objectivity may have been lost. As a result, all of the rejections are improper and are appealed.

V. In some instances Applicant has not had a meaningful opportunity to advance prosecution on the merits due to rejections that have been somewhat difficult to understand.

7. Argument

I. Claims 1, and 4-13 were rejected under 35 U.S.C. §102(e) as being anticipated by Elworthy (US 6,125,362)

Claims 1 and 4-6

Independent Claim 1 was rejected under 35 U.S.C. §102(e) as being anticipated by Elworthy. Elworthy does not teach each and every element of claim 1 and thus fails to support the §102 rejection. Therefore, the rejection should be withdrawn. Additionally, the rejections of dependent claims 4-6 should be removed.

The claimed system provides an elegant and efficient technique for determining a document language. The elegance and efficiency come from the setting of a negative assumption for a language. Elworthy does not teach a logic for setting a negative assumption value for each of the candidate languages. Elworthy only provides an accumulator that is initialized to zero and that is used to count probabilities for all languages. The claimed system facilitates reducing the size of a problem space based on the negative assumption values for candidate languages. In Elworthy, all candidate languages are processed for each word without considering negative assumption values.

The negative assumption value is processed by a "language analyzer" recited in claim 1. The language analyzer retrieves probability values from a database and adjusts the negative assumption until a language is determined. This feature is also not disclosed by Elworthy, which never actually selects a language.

Elworthy suggests a most likely possible language for a document by accumulating probabilities associated with each token for each language for an input data (e.g., document). Elworthy compares the final accumulated totals to see which language has the highest total. Nowhere are any negative hypotheses established and/or proven. If any were, then Elworthy would describe processing for selecting and/or eliminating a language before 100% of the terms had been processed through 100% of the candidate languages. In summary, Elworthy counts up scores until all of the input data has been processed and compares the total scores without "proving or disproving assumptions". In Elworthy, the winning score may even be the initial score. This makes the "negative

assumption" proposed by the Office Actions act as the positive assumption at the same time. Since a thing can not be and not be at the same time from the same point of view, logic requires dismissal of this proposal in the Office Actions.

In the claimed system, it is possible that not all languages will be processed to completion. For example, the present specification shows example results in Table 1 on page 12 where after 2 iterations the candidate language of English can be eliminated as a possibility since the negative hypothesis has been proved. The negative hypothesis is proven because enough words having such a high probability of not being English have been encountered. As a result of proving the negative hypothesis the problem space is reduced and English is no longer considered. In Elworthy, no such reduction is possible and thus it will process all input data for every language to determine who has the highest accumulated probability. The Office Actions recite that they can not find this pruning and dismissal in the claims. However, the operations of the language analyzer with respect to adjusting the negative assumption value illustrates these actions. This rejection is akin to a rejection of arguments concerning a claim that describes a process for combining two hydrogen atoms and one oxygen atom (H_2O) where the arguments mention water. While the word water may not appear in the claim, arguments that use the term water should not be rejected out of hand.

Elworthy does not teach a database having text strings each having an associated probability value that indicates a probability that the text string occurs within a language, where the probability is based on occurrences of the text string in all candidate languages. In Elworthy, the probabilities associated with a text string are based on the probability that the text string is part of a single language.

The Office Actions cite Elworthy (column 7, lines 50-65) as teaching the claimed probability limitation. However, the cited section of Elworthy, shows no such thing. Phrase by phrase analysis reveals that the limitation is not present.

Elworthy Phrase	Probability Based On Occurrences Of The Text String In All Candidate Languages Disclosed?
Methods which can be used are the methods described in the articles by Sibun & Spitz and Sibun & Reynar.	No
The word tokens are then input to each of the lexicons 25a,25b,25c . . . 25L for the languages to which the OCR data may belong.	No
The lexicons 25a,25b,25c . . . 25L comprise predetermined probability values that the word token belongs to the language.	No
The probability output from the lexicons 25a,25b,25c . . . 25L are input to respective accumulators 26a,26b,26c . . . 26L where the probabilities for sequential word tokens are accumulated to form an accumulated probability.	No
The accumulated probabilities of each of the accumulators 26a,26b,26c . . . 26L are input to a comparator 26 wherein the probabilities are compared with one another and with a predetermined threshold to determine whether a language is uniquely identifiable as the language to which the OCR data belongs.	No

In this passage Elworthy teaches that its lexicons comprise "predetermined probability values that the word token belongs to the specific language", not that the probability is "based on occurrences of the text string in all of the candidate languages" as claimed.

In summary, claim 1 recites features not taught or suggest by Elworthy. Thus Elworthy fails to support the §102 rejection and the rejection should be withdrawn. Claim 1 therefore patentably distinguishes over the references of record and is in condition for allowance. Accordingly, dependent claims 4-6 also patentably distinguish over the references and are in condition for allowance.

Claims 7, and 8-13

Independent Claim 7 was rejected under 35 U.S.C. §102(e) as being anticipated by Elworthy. Claim 7 describes a method that includes setting a null hypothesis,

determining a contrary probability, adjusting the null hypothesis, and determining the document is in one language. Elworthy does not teach any of these elements and thus fails to support the §102 rejection. Therefore, the rejection should be withdrawn. Additionally, the rejections of dependent claims 8-13 should be removed.

Claim 7 describes "setting a null hypothesis to a true value...". The Office Action cites Elworthy column 12, lines 20-38 and claim 13 as teaching this element. Elworthy states, "the accumulator is initially zeroed..." (column 12, line 22). The Office Action equates zeroing out a counter to setting a null hypothesis to a true value. This is mathematically and logically inaccurate because the accumulator is not used to prove or disprove a hypothesis. If it were, then a test on the value of the accumulator would occur at some point, maybe even on each iteration, and a language could be selectively dropped from further consideration and/or selected based on the comparison. This does not happen. In Elworthy, probabilities are added to the initially zeroed out accumulator to compute a total value. When all words have been processed for all languages, the totals in all the accumulators are compared. However, the values are only compared to see which has the highest value and whether the highest value exceeds a threshold. The accumulator may be used to suggest a result, but no test is ever performed to prove that a language is no longer a candidate (e.g., a negative hypothesis). The lower valued accumulators are essentially ignored. An ignored item neither proves nor disproves anything.

The examination insists that proving a positive disproves a negative. This is simply not the case in related event statistics. In simple statistics, proving that a die roll is a 6 proves that it is not a 5. However, in more complicated statistics, proving that a card is an Ace does not prove that it is not a spade.

Elworthy fails to teach "determining a contrary probability..." as recited in claim 7. Elworthy produces only positive probabilities. These positive probabilities do not disclose determining a contrary probability at all, let alone a contrary probability "based on probabilities that the text string belongs to each of the candidate languages" as recited in claim 7. Thus, additionally, Elworthy does not disclose that the probabilities are "based on occurrences of the text string in all of the candidate languages." Instead

Elworthy determines term probabilities based on probabilistic models that are "independent of the others" (Col. 8, lines 5-7).

Regarding the limitation of "determining the document is one language...", the Final Office Action cites Elworthy column 12, lines 20-38 and column 13, lines 22-35, and reasons that "the highest accumulated probability-accounts for approval and simultaneously disproval, C. 13, lines 44-58." However, determining that English is the most likely language does not include proving that French cannot be the language. All Elworthy provides is a suggestion that one language is more likely the correct language than another. The "overlap" condition described below shows the depth of the incorrectness of the Examination. Elworthy explains:

> In FIG. 14b it can be seen that the probability of the language being English has exceeded the threshold but there is still overlap with the probability for the languages being French and Italian. If there is no more data these **three languages could be identified as possible languages** to which the input data belongs. (Elworthy, column 13, lines 49-54, and Figure 14b) [emphasis added]

Thus, even passing the threshold and having the highest score does not prove or disprove anything, it simply makes one thing more likely than another. In the example, the result is that English scored high enough that it is likely the language but French and Italian also are good choices. No determination is made, leaving Elworthy void of the teaching that a "determining" element includes disproving the null hypothesis by approaching the false value.

Based on the above explanations, Elworthy fails to teach each and every feature of claim 7. Thus, Elworthy fails to support a proper §102 rejection and the rejection should be withdrawn. As such, claim 7 patentably distinguishes over the references of record and is in condition for allowance. Accordingly, dependent claims 8-14 also patentably distinguish over the references and are in condition for allowance.

Dependent Claim 10

Claim 10 depends from claim 7 which has been shown to be not anticipated and thus claim 10 is similarly not anticipated. Additionally, claim 10 recites that the claimed

method includes pregenerating probability data corresponding to each candidate language. This probability data includes a probability value for a text string that is normalized based on an occurrence probability of the text string in all the candidate languages. In Elworthy, to the extent that probabilities are computed, they are not computed in this way (e.g., normalized).

The Final Office Action (page 6) cites Elworthy column 2, lines 30-38 as teaching the claimed normalization. However, this passages describes comparing probability values and does not teach normalization as claimed. Normalizing a probability value as recited in claim 10 involves changing the values of the probability values. This is understood by one of ordinary skill in the art. The "comparing" of values described in Elworthy does not result in the changing of values. Indeed, one of ordinary skill in the art understands that "normalizing" is not "comparing" and that "comparing" does not teach "normalizing".

Elworthy does not mention normalizing or any form of normalization in its disclosure. That is because Elworthy does not perform the normalization involved in related item processing. Normalization to account for cross language relations is not needed in Elworthy because in Elworthy the term probabilities are computed under the rule that "the probabilistic model for one language is independent of the others." (Elworthy, column 8, lines 5-7). This independence obviates the need for normalization techniques. Thus the method of claim 10 is not taught or suggested by Elworthy and the rejection should be withdrawn.

II. Claims 15, 16, 20, and 21 were rejected under 35 U.S.C. §102(b) as being anticipated by Pon et al. (US 6,047,251)

Claim 15

Claim 15 concerns a process for determining that a document is in a selected language. The process includes setting a probability assumption that indicates that the document is not in the selected language. Pon includes no such setting. Like Elworthy, a score is initially zeroed. However, this zero score could end up as the highest score for a language and thus indicate that the language is the most likely language. A value that indicates that a language is the most likely language does not anticipate a probability assumption that a language is not a selected language.

Additionally, claim 15 describes the process including disproving the probability assumption based on a contrary probability. Pon, like Elworthy, only processes positive probabilities, not contrary probabilities. Pon discloses an optical character recognition system that uses a dictionary-based approach to identify languages in a document (see Abstract). Pon is a stripped-down version of Elworthy. Instead of accumulating the complex probabilities that Elworthy generates, Pon counts the number words from a document that matches words in a dictionary for a specific language.

> the confidence statistic can be computed by counting the number of words in the zone that are found in each of the respective dictionaries.
> (Pon, column 5, lines 63-65)

> The language with the highest confidence statistic is ascertained, and used as an initial estimate of the language for the zone.
> (Pon, column 6, lines 1-3).

The "confidence statistic" as described by Pon is an accumulated total of the number of words in a document region that are found in a dictionary. Basically, Pon adds a "1" to a counter for each word match (Pon, column 7, lines 1-3) and "adds" a zero when there is no match. Adding a zero does not teach adjusting a negative assumption or manipulating a null hypothesis, since adding zero has no effect whatsoever. When processing is finished, Pon finds the highest score (Pon, column 8, lines 4-9). Conceivably the highest score could be the initial zero.

Similar to Elworthy, Pon neither "proves" nor "disproves" any assumption but accumulates points and takes the highest score. At no point can a language be removed from consideration and only at the end can a language be recommended. In both cases the entire problem space is analyzed with no pruning possible because there are no hypotheses being tested that would permit the removal of any language from consideration.

Claim 15 also recites "if the contrary probability fails to support the probability assumption, then the document is determined as being in the selected language." Applicant finds no teaching in Pon where any value is used to perform this process. Like Elworthy, the only values that are processed are the accumulators that are examined to determine who has the highest score.

For the reasons set forth above, a proper §102 rejection of claim 15 has not been established and the rejection should therefore be withdrawn. Thus claim 15 is now in condition for allowance. Accordingly, dependent claims 16-21 are also in condition for allowance.

III. Claims 14, and 17-19 were rejected under 35 U.S.C. §103(a) as being unpatentable over Pon in view of Elworthy

Claim 14 depends from claim 7 which has been proven to be not anticipated and thus this claim cannot be obvious. Additionally claim 14 recites how the contrary probability is determined. The determining includes normalizing a sum of occurrences of a string found in a sample set of documents from all the candidate languages. Neither Elworthy nor Pon disclose this normalizing. For this additional reason this claim is not obvious and is in condition for allowance.

Claims 17-19 depend indirectly from claim 15 which has been shown to be not anticipated and thus these claims can not be obvious. Additionally, each of these claims recite additional elements concerning generating a probability database. Neither Elworthy nor Pon describe generating the database in the manner described. By way of illustration, with respect to claim 17, neither reference describes producing a contrary probability where the contrary probability is based on an occurrence frequency of a string in one language as influenced by a total occurrence frequency of the string in all the candidate languages. By way of further illustration, with respect to claim 18, neither reference describes determining the occurrence frequency based on a sample set of documents. By way of further illustration, with respect to claim 19, neither reference describes normalizing the contrary probability. For at least these additional reasons these claims are not obvious and are in condition for allowance.

IV. MPEP §2141.03 requires that Office Actions ascertain and describe the level of the hypothetical person of ordinary skill in the art so that objectivity can be maintained. Here the Office Actions neither ascertained nor reported on the level of ordinary skill in the art. Thus, all the rejections are improper and are appealed.

The danger inherent in not understanding the level of skill of the hypothetical person of ordinary skill is very evident in this case. One of ordinary skill would understand not just simple statistics (e.g., die rolls), but also conditional probabilities concerning related events and/or items. One of ordinary skill would also understand the difference between the "last man standing" class of algorithms and the "rank all contenders" class of algorithms.

In the Advisory Action, and throughout the office actions, the examination has insisted that the Elworthy probabilities anticipate the claimed probabilities. However, Elworthy states that the "probabilistic model for one language is independent of the others." (Col. 8, lines 5-7). Since the probabilistic model for each language is admittedly independent, then the individual probability computed for any term with respect to a language using that independent model must by definition not consider occurrences across all languages. One of ordinary skill in the art would appreciate this.

The claimed probabilities are not built on these independent models. The individual probability computed for any term is "based on occurrences of the text string in all of the candidate languages." A probability that depends on occurrences in all languages is fundamentally different from a probability computed from a probabilistic model that is "independent of the others." This fundamental difference allows the claimed negative assumption processing, which is absent in both Elworthy and Pon.

The examination has also insisted that setting an accumulator to zero teaches a logic for setting a negative assumption value. The examination asserts that the accumulated probabilities concerning a language "inherently determines the value that a character string does not belong." Advisory Action, Page 2, last line. One skilled in the art would not make this mistake. The reasoning is flawed on both logical and mathematical grounds. The accumulation of positive probabilities in Elworthy and Pon describes how likely a language is the "correct" language. Let this value be X. The

Examination insists that (1 – X) must therefore inherently represent the likelihood that the language is not the correct language.

This misunderstanding may be attributed to not determining what would be understood to one skilled in the art. The misunderstanding arises from applying the rudimentary statistics of independent events (e.g., the statistics of dice) to the more complicated problem of conditional probabilities involving non-independent events (e.g., the statistics of multiple table poker). In the rudimentary statistics applied, the likelihood that a six sided die will roll a 6 equals 1 in 6, which is one minus the sum of the probabilities that the six sided die will roll a 1, 2, 3, 4, or 5. This rudimentary statistical analysis is evidenced when the Final Office Action asserts "[i]t is inherent to a positive step of proving a probability assumption, that disproving a probability assumption is also realized." Final Office Action, Page 2, Paragraph 2. While this may be true in elementary statistics, it is not necessarily true in more advanced statistics.

The claimed probability for a term being computed in light of a term's presence in all candidate languages does not yield a simple independent result. For example, a term may appear in more than one language and the probability will be computed based on considering all the languages. For example, assume that the term "auto" appears in at least English and French, and does not appear in Serbian. In Elworthy, the presence of this term would add to each of the English and French accumulator and would not affect the Serbian accumulator. Conversely, in the invention, the probability would affect the negative assumption for each language, perhaps strengthening the negative assumption for Serbian so much that it is removed as a candidate language. In Elworthy, the amount added to the French accumulator would be determined by the probability that "auto" appears in the French language, where the probability was determined by a probabilistic model "independent from the others" (e.g., English, Serbian). Similarly, the amount added to the English accumulator would be determined by the probability that "auto" appears in the English language, once again where the probability was determined by a probabilistic model "independent from the others" (e.g., French, Serbian). Adding these values to the English and French accumulators may indicate that English or French is more likely the language but, unlike the claimed invention, adding these values indicates nothing about how likely it is that Serbian is not the language. The Serbian counter likely

remains unchanged in this scenario since no contrary probability associated with "auto" with respect to Serbian is processed. In the claimed invention, the presence of the term "auto" in the document for which the language is being determined would have different consequences. One skilled in the art would appreciate this.

In conclusion, one skilled in the art would understand that in the claimed invention, each term has a probability that facilitates answering the question, "does this term allow me to rule out any languages." One skilled in the art would also understand that in the reference each term has a probability that only facilitates answering the question, "how likely is it that this term is a member of this language?" One skilled in the art would appreciate that these are fundamentally different systems and would allow the claims.

V. In some instances Applicant has not had a meaningful opportunity to advance prosecution on the merits due to rejections that have been somewhat difficult to understand.

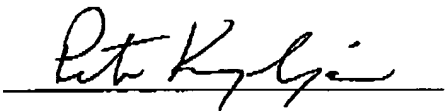
In some cases Applicant has been unsure of the specific point being raised in a rejection and thus has been prejudiced in advancing prosecution. Consider, for example, the following statement that was provided in the Advisory Action on page three at lines 1-2. **"Claim 7 to claim 1, and thus the arguments are not persuasive, wherein contrary probability, see as the negative assumption value."** Applicant can not advance meaningful prosecution on the merits concerning this rejection. For this additional reason the rejections are improper and should be reversed.

Conclusion

For the reasons set forth above, prima facie §102 and §103 rejections have not been established for any claim. Thus, all rejections are improper and should be reversed. Accordingly, claims 1, 4-21 patentably and unobviously distinguish over the references of record and are now in condition for allowance. An early allowance of all claims is earnestly solicited.

Respectfully submitted,

Aug. 14, 2006
Date



Petar Kraguljac (Reg. No. 38,520)
(216) 348-5843

John. T. Kalnay (Reg. No. 46,816)
(216) 348-5844

McDonald Hopkins Co., LPA
600 Superior Avenue, E.
Suite 2100
Cleveland, OH 44114

Claims Appendix

1. A system for automatically determining a language of a document from a set of candidate languages, the system comprising:

a database containing probability data for a plurality of text strings each having a predetermined length equal to each other, each text string of the plurality of text strings having an associated probability value indicating a probability that the text string occurs within a language based on occurrences of the text string in all of the candidate languages;

logic for setting a negative assumption value for each of the candidate languages indicating the document is not one of the candidate languages;

an extractor for extracting a character string from the document, the character string having a length equal to the predetermined length of the plurality of text strings contained in the database; and

a language analyzer for determining a probability value for each of the candidate languages that the character string does not belong to the candidate languages by retrieving the probability value associated to the character string from the database for each or the candidate languages, and includes logic for adjusting the negative assumption value based on the probability value, the language analyzer determining that the document is one language of the candidate languages when the negative assumption value passes a threshold value.

2. (Canceled)

3. (Canceled)

4. The system as set forth in claim 1 further including an information retrieval engine for retrieving documents in response to a search request, the documents retrieved being analyzed by the language analyzer.

5. The system as set forth in claim 1 wherein the logic for adjusting includes logic for combining the negative assumption value with the probability value.

6. The system as set forth in claim 1 wherein the language analyzer further includes iteration logic for causing the extractor to extract another character string from the document if the negative assumption value fails to pass the threshold value.

7. A method of determining a language of a document from a set of candidate languages, the method comprising the steps of:

setting a null hypothesis to a true value for each candidate language indicating the document is not in the candidate language and setting a false value;

extracting a text string from the document, the text string having a predetermined length;

determining a contrary probability for each candidate language that the text string does not belong to the candidate language based on probabilities that the text string belongs to each of the candidate languages where the probabilities are retrieved from a database that stores probability values for a plurality of text strings each having the predetermined length, each text string of the plurality of text strings having an associated probability value for each candidate language indicating a probability that the text string occurs within a language from the candidate languages based on occurrences of the text string in all of the candidate languages;

adjusting the null hypothesis for each candidate language with the contrary probability corresponding to the candidate language; and

determining the document is one language from the candidate languages when the null hypothesis for the one language is disproved by approaching the false value.

8. The method as set forth in claim 7 further includes setting a threshold value indicating that the null hypothesis is disproved.

9. The method as set forth in claim 8 further includes repeating the extracting step for a different text string from the document and repeating the method until the null hypothesis is disproved for one of the candidate languages by passing the threshold value.

10. The method as set forth in claim 7 further includes pregenerating probability data corresponding to each candidate language, the probability data including a probability value for a text string that is normalized based on an occurrence probability of the text string in all the candidate languages.

11. The method as set forth in claim 7 further includes identifying the document based on a search request.

12. The method as set forth in claim 7 wherein the extracting step includes extracting a plurality of sequential characters that form the text string.

13. The method as set forth in claim 7 wherein the setting step includes setting the true value to 1 and setting the false value to 0.

14. The method as set forth in claim 7 wherein the contrary probability for a first candidate language is determined based on a number of occurrences of the text string found in a sample set of documents from the first candidate language which is normalized by a sum of occurrences of the text string found in a sample set of documents from all the candidate languages.

15. A process of determining that a document is in a selected language, the process comprising the steps of:

setting a probability assumption indicating that the document is not in the selected language;

extracting a character string from the document; and

disproving the probability assumption based on a contrary probability that the character string does not belong to the selected language such that if the contrary

probability fails to support the probability assumption, then the document is determined as being in the selected language.

16. The process as set forth in claim 15 further includes determining the document is the selected language from a set of candidate languages.

17. The process as set forth in claim 16 further including generating a probability database having a contrary probability for each of a plurality of character strings for each of the candidate languages, where the contrary probability of a character string in one language is determined based on an occurrence frequency of the character string in the one language influenced by a total occurrence frequency of the character string in all the candidate languages.

18. The process as set forth in claim 17 further including determining the occurrence frequency of each character string based on a sample set of documents provided for each of the candidate languages.

19. The process as set forth in claim 17 wherein the contrary probability of the character string in one language is normalized by the total occurrence frequency of the character string in all the candidate languages.

20. The process as set forth in claim 15 further including identifying the document in response to a search request.

21. A computer program product configured to perform the process of claim 15.

Evidence Appendix

There is no extrinsic evidence.

Related Proceedings Appendix

There are no related proceedings.